



Folha de Exercícios n.º PL04	
Unidade Curricular:	Aplicações Informáticas na Biomedicina
Ano Letivo:	2019/2020 - 1.º Semestre

1 Criação e Execução da Primeira Job no Talend

O ficheiro disponibilizado juntamente com esta ficha prática laboratorial, nomeadamente “world_happiness_2016.csv”, contém os dados reais relativos à classificação da felicidade por país em 2016 de acordo com diversas variáveis.

O *dataset* é constituído por 13 diferentes colunas, nomeadamente *country*, *region*, *happiness_rank*, *happiness_score* (de 0 a 10), *lower_confidence_interval*, *upper_confidence_interval*, *economy*, *family*, *health* (esperança de vida), *freedom*, *trust* (corrupção governamental), *generosity* e *dystopia_residual*.

É importante referir que as colunas *economy*, *family*, *health* (esperança de vida), *freedom*, *trust* (corrupção governamental), *generosity* e *dystopia_residual* descrevem em que medida esses fatores contribuíram para avaliar a felicidade em cada país.

Tendo em conta o seu enquadramento teórico e prático sobre o processo de ETL e o Talend, os objetivos principais a serem alcançados com a resolução desta ficha prática laboratorial são a sua ambientação à ferramenta *Talend Open Studio (TOS) for Big Data*, bem como a criação e a execução da sua primeira *job* com as componentes disponibilizadas na mesma.

Com esta ficha prática laboratorial, pretende-se que:

1. Crie um novo *schema* no MySQL Workbench denominado “Ficha4”.
2. Abra o Talend e crie uma *job*, recorrendo às componentes adequadas, que permita, de forma geral:
 - (a) Conectar-se à base de dados criada no passo anterior (*tDBConnection*).
 - (b) Fazer o *import* dos dados no ficheiro denominado “world_happiness_2016.csv” (*tFileInputDelimited*).
 - (c) Fazer o *export* de toda a informação no ficheiro world_happiness_2016.csv para uma nova tabela denominada “world_happiness_2016” no *schema* Ficha4 (*tMap* e *tDBOutput*).
 - (d) Fazer o *export* dos dados na tabela world_happiness_2016 de todos os países em regiões europeias para uma nova tabela denominada “world_happiness_2016_europe” no *schema* Ficha4 (*tFilterRow* e *tDBOutput*).
 - (e) Criar um ficheiro Excel (.xls), nomeadamente world_happiness_2016_europe_good.xls, com os dados relativos aos países europeus que tenham um *happiness_score* superior ou igual a 5 (*tFilterRow* e *tFileOutputExcel*).
3. Corra a *job* criada.